Executive Summary: Chicago West Nile Virus Action Plan

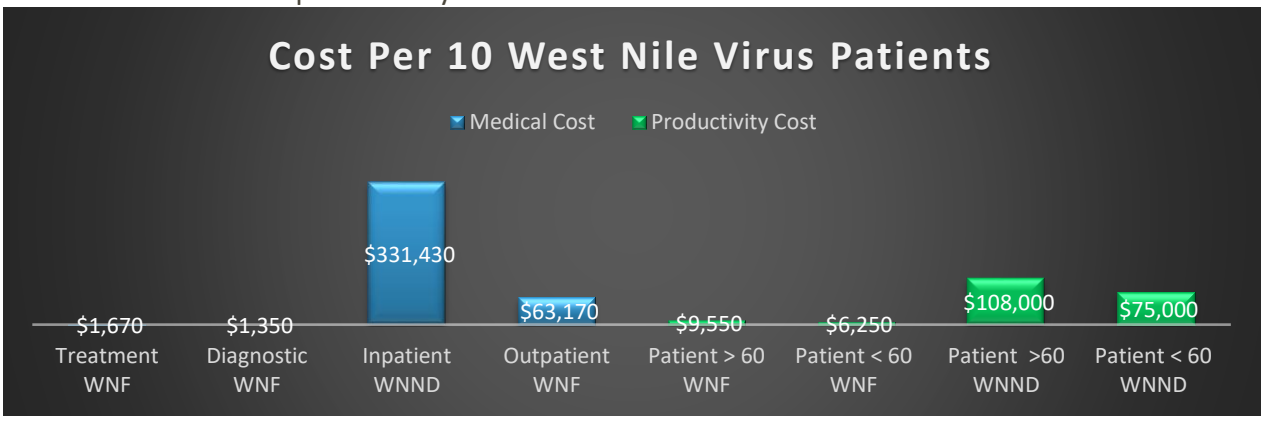
SMARRT Consulting Group

Problem Statement

In September 2001, West Nile virus was first identified in Illinois when laboratory tests confirmed its presence in two dead crows discovered in the Chicago area. This comes only two years after West Nile virus first emerged in the United States in New York in the fall of 1999. By the end of 2002, Illinois had counted more human cases (884) and deaths (64) than any other state in the United States.

This is where SMARRT consulting group can help to prevent this costly pandemic from resurfacing and preserve public safety. SMARRT Analytics has focused exclusively on consultation in matters of public health to assess the risk of disease outbreaks in cities across the world to provide analytical expertise ultimately providing recommendations for intervention and prescriptive prevention.

In 2002, West Nile virus was discovered in Chicago for the first time with over 225 cases reported. In response, the Chicago Department of Public Health (CDPH) has implemented a city-wide surveillance and mosquito control measure program. Based upon a 2010 study produced by the CDC, a 2005 outbreak of West Nile virus cost Sacramento County, California $2.98 million. West Nile Virus (WNV) can have two different effects on a human host, West Nile Fever (WNF) and it's much more severe and costly West Nile neuroinvasive disease (WNND). A cost benefit analysis was also performed during this study, which indicated that preventative measures such as spraying would only need to prevent 15 cases of WNND to make the control measure cost effective.[3] Today the population of Chicago is 2.7 million which is 1.8 times the size of Sacramento County, California and the outbreak we experienced in 2002 having 225 infected compared to the 163 cases in California would result in an overall cost of over $4 million in medical and productivity costs alone.



Goals

SMARRT Consulting Group will help Chicago:

1. Provide a final optimized regression model to predict the number of mosquitos in an area and a classification model to predict the presence or absence of West Nile Virus (WNV).

2. Identify the highest risk areas using these models to provide prophylactic spraying recommendations.

3. Deliver a real time dashboard and application to show the highest risk areas based on likelihood of occurrence (LoO) and severity of effect (SOE).

Technology Assessment

SMARRT Analytics will use the following toolkit to complete the project:

- CRAN R 3.3+
- Python 3.5
- RStudio
- Jupyter-Lab
- Tableau Desktop
- Microsoft Excel
- SAS JMP Pro 13
- ESRI ArcGIS
- PlantUML

The primary analytical tool used is the open source statistical software R. The ability to customize models specifically for WNV and tuned to account for the costs of inaccurate predictions, make R an obvious choice. Several factors point to the use of R:

- It is free, the lowest cost of any available tools
- The EDA, visualization, and breadth of model-building packages ranks very high
- The Shiny package allows for interactive visualizations to integrate new predictions

Prediction & Classification Tools/Approach

SMARRT Consulting Group has utilized a 2-tiered approach towards developing predictive models, since there are two important measured to be tracked:

1. The number of mosquitoes at different locations over time
2. If the mosquitoes will have the West Nile Virus

In order to predict **when** and **where** mosquito populations will grow and be susceptible to spreading West Nile Virus to the surrounding human population, we needed to use multiple data sources that provide insights into the time- and location-specific contexts where mosquitoes carrying West Nile Virus are seen. To that end, we deployed cutting-edge models. These modeling methods made use of daily weather data and neighborhood characteristics that change over time, including demographic and socio-economic status indicators from the Census' American Community Survey and neighborhood building violations and vacancies from the City of Chicago tied to specific time periods and areas, which is a way to identify times and places where stagnant water and other conditions suitable to mosquitoes are likely to exist.

In order to parse through these elements and determine what data are most helpful for a particular time and place, we used sophisticated "feature selection," "dimension reduction" and "cross validation" techniques to develop the models that perform best for predicting data the models have not yet seen -- in other words, to make predictions similar to what will be done in the real world to guide mosquito abatement and West Nile Virus control efforts. These models included statistical approaches like linear regression and machine learning approaches like random forests to estimate how many mosquitoes will be present, and machine learning approaches like extreme gradient boosted forests to determine how likely they were to have West Nile Virus. Model performance was quantified and tested rigorously using the aforementioned cross validation techniques.

At this stage of the modeling, the models show quite decent performance. The team has identified steps to improve the performance of the model, including methods like tuning of configurable parameters, selection of various thresholds. We will also incorporate calculations of the financial impact of the different types of errors. We will tune the models to minimize this financial impact to the city.

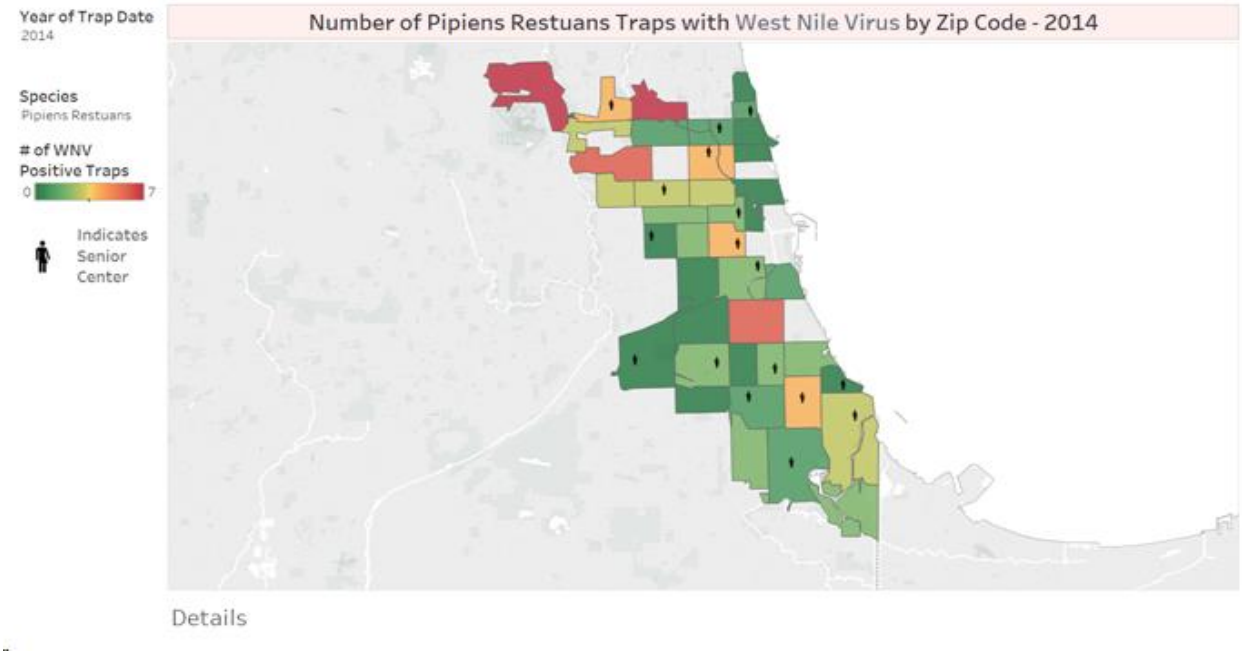
Dashboard Development

Maps are a powerful tool that can be used to effectively visualize geospatial data. SMARRT Consulting Group will deliver several maps to support the findings from our analysis. The maps will be incorporated into different Tableau dashboards that will be interactive. This will enable the CDPH to do some self-service analytics within the environment we create.

The main type of map we plan to deliver are heat maps. Heat maps are a graphical representation of data that transform the quantitative data into color. Combining the color dimension with the geospatial will allow very simple ease of use for the CDPH.

Other aspects we will incorporate into our dashboards are markers to indicate high risk areas. We are defining high risk areas as those where a large outbreak of West Nile virus infections would have the worst impact due to a more sensitive population group. This would include areas with a lot of schools or daycare centers or areas with retirement homes.

We currently have three separate dashboards, each with a different objective. The first is just a simple look to show over the 10-year time frame (2007-2016) the total number of mosquitoes vs. the number of positive WNV traps by species. The species can be selected, and the dashboard will update. The second dashboard shows the number of mosquitoes aggregated at a zip code level in a heat map. This map is also able to filter specifically on a certain mosquito species or to look at all species together. You also have the functionality to change what year you are looking at. We have been able to incorporate the senior center locations onto this map, notated by a person pictorial. The third dashboard has the same capabilities and features of the second dashboard but is instead showing the number of positive WNV traps by zip code.



App Development

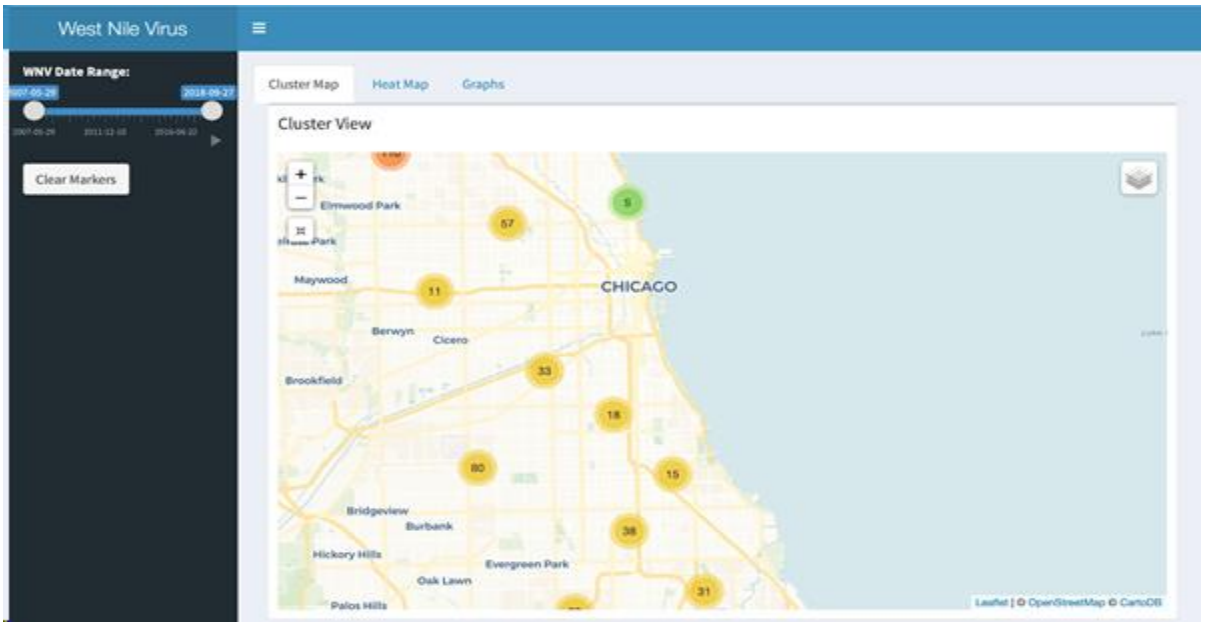
The primary objective of SMARRT Consulting Group is to decrease the amount of West Nile virus infections in a cost-effective way. According to the Center for Disease Control and Prevention (CDC) “The most effective way to prevent infection from West Nile virus is to prevent mosquito bites.” By providing accurate and real-time forecasts on high risk areas for testing positive for WNV, the public can take preventative measure against infection from

WNV. SMARRT Consulting Group will deliver the WNV forecasts to the public through a dashboard on a public website as well as through a free mobile application.

WNV forecasts will be determined through two methodologies. A series of models will determine the likelihood of an area to test positive for WNV. The severity will be identified through a red, amber, or green monitoring system (RAG). The RAG status will be determined by calculating the proximity of forecasted WNV to a school, senior center, or hospital. These establishments are used since children, the elderly, or people with weakened immune systems are more likely to contract WNV.

The data engineering team processed several WNV files and produced a WNV master file. This is the primary file used in the applications. Additional input files include a Chicago shapefile and several files which include locations of Chicago hospitals, schools, and senior centers. The modelling data will be integrated into the master WNV file for ease of use.

The mobile applications are divided into three sections: Cluster, Heat-map, and Graph tabs. The clusters tab shows a dynamic map of Chicago and provides longitude and latitude coordinates of historical and forecasted locations for positive cases of WNV. The heat-map of Chicago overlays the severity for WNV; red areas indicate a high severity and green indicating low severity for WNV. The graph tab shows key WNV metrics such as total cases identified, current forecast (RAG), and areas most likely to test positive for WNV.



Project Status

Project Activity and Deliverables	Target Completion	Status
Team Selection	Week 2	Complete
Define team roles	Week 2	Complete
Project Definition	Week 3	Complete
Project Goals Assignment	Week 3	Complete
Exploratory Data Analysis	Week 4	Complete
Model Build and Refinement	Week 8	On Schedule
Initial Findings/Exec Summary	Week 6	Complete
Model Selection/Recommendations	Week 7	On Schedule
Dashboard/App Design	Week 8	Ahead of Schedule
Final Report	Week 9	On Schedule
Final Presentation	Week 10	On Schedule

Next Steps

- Continue model refinement and test final models on test data
- Select best regression and classification model for integration
- Integrate both models into dashboard and application
- Gather and document all recommendations into Final report
- Develop Final presentation

Preliminary Conclusions

Modeling

SMARRT Consulting has demonstrated that we can effectively identify, gather and combine disparate data sources to make effective predictive models for identifying temporally and geospatially specific mosquito and WNV risk.

Dashboard

SMARRT Consulting has demonstrated the ability to create various interactive dashboards with drill down capabilities to help provide self-service analytics solutions for CDPH to geospatially explore the number of mosquitoes and the number of positive WNV traps. These findings can be explored by mosquito species and by year.

Shiny and Mobile Application

SMARRT Consulting has demonstrated the ability to render the modelled WNV data in an “easy to understand” format and provide targeted alerts the public.